OGC Data Service Catalog

Allan Doyle
International Interfaces, Inc.
946 Great Plain Ave. PMB-182
Needham, MA 02492

Yonsook Enloe
SGT Inc.
7701 Greenbelt Rd
Greenbelt, MD 20770

Ananth Rao
SGT Inc.
7701 Greenbelt Rd
Greenbelt, MD 20770

***Abstract*- The motivation behind the Open GIS Consortium (OGC) Web Services service model work is described. Historical use of data catalogs in the Earth Observation (EO) community has influenced the current OGC Catalog Services specification, which in turn is being adapted to provide service catalog functionality. A prototype implementation is described, followed by discussion on technical issues that remain as well as future work within the OGC/EO community.**

I. INTRODUCTION

The Earth observation information systems developed in the past 10 years concentrated on basic data metadata search, simple visualization, and data order. The Earth Observing System Data and Information System (EOSDIS) Version 0 (V0) catalogs (operational since 1994) and the EOSDIS Core System catalogs (operational since 1998) are good examples of data metadata search and order systems. Many other international space agencies (e.g. the European Space Agency, the Canadian Center for Remote Sensing, and the National Space and Development Agency of Japan) operate similar Earth observation catalog systems. Studies of global change necessarily involve data from multiple instruments, covering many parts of the Earth, with data sources from multiple agencies. International cooperation and collaboration for Earth observation catalog systems benefits the science community by facilitating common access to multiple Earth observation data sources. Internationally developed standards for catalog interfaces can provide common methods to access the Earth observation catalog systems.

This paper discusses the evolution of catalogs from purely data-oriented to service-oriented and mixed data/service oriented content, beginning with the Committee on Earth Observing Satellites' (CEOS) Catalog Interoperability Protocol (CIP) - the first Earth observation specific catalog standard developed by a consortium of CEOS agencies in 1995-1997. In 1997-1998, CIP was taken beyond the CEOS community and with some modifications was made into an international catalog interface standard at the Open GIS[1] Consortium (OGC). The concepts in the OGC Catalog Interface standard were extended beyond data metadata catalogs to data service catalogs during the OGC Interoperability Program Web Mapping Testbed (WMT Phase 1 in 2000). This paper discusses the concepts that have led to a Data Services Catalog. A simple prototype that implements the OGC Catalog Interface and a data service interface is then discussed. The paper concludes with a look into future work and a discussion of outstanding issues.

II. CIP

Members of the Committee on Earth Observation Satellites (CEOS), a consortium of international agencies that operate Earth observation missions, have collaborated on catalog techniques since the early 1990s. The result of the collaboration produced the CEOS Catalog Interoperability Protocol (CIP), an Earth observation customization of Z39.50 (a search and retrieval protocol used by the library community for their catalog information systems).

The following themes were used to design CIP:

- A hierarchical collections concept for distributed searching
- Middleware to route messages
- The Data Model defined was based on existing standards such as FGDC[2], GCMD DIF[3], and the EOSDIS Core Metadata
- Secure ordering was specified as a Z39.50 Extended Service

CIP provides standardization for search, retrieval, order procedures, and metadata attributes. A National Aeronautics and Space Administration (NASA) engineering team was part of the international team of engineers who designed and specified CIP. As a result, CIP contains many EOSDIS Core System concepts. Much of the EOSDIS Core metadata is replicated in the CIP metadata. EOSDIS catalog capabilities for directory, inventory, browse, and order functions are also found in CIP.

[1] "GIS" itself is an acronym meaning Geographic Information System. The OGC is an industry trade association of approximately 200 international government, university, and commercial member organizations working on interoperable interfaces for spatial information processing. See http://www.opengis.org for more details.

[2] Federal Geographic Data Committee "Geo" protocol

[3] Global Change Master Directory - Directory Interchange Format

III. OGC CATALOG INTERFACE

In 1997, the Open GIS Consortium, an international standards organization for geospatial information systems, composed of representatives from government agencies and from commercial companies implementing the standards in commercial products, issued a Request For Proposals (RFP) for a Catalog Interface. The same NASA engineering team that helped define CIP worked with representatives from other government agencies and commercial companies to define an OGC Catalog Interface. This interface was approved as an OGC Catalog Services Specification [1] in August 1998. The OGC Catalog Services Specification is an improved, updated, and simplified version of CIP.

The OGC Catalog Interface is a message based, stateful interface containing constructs for the mandatory functions of search and retrieval, and for the optional functions of data access (order) and management. The Z39.50 ASN.1[4] messages have been transformed into eXtensible Markup Language (XML) [2] based messages for easier implementation. It supports multiple query languages and is data model independent. Several vendors have implemented the OGC Catalog Interface in commercial products.

As the OGC Catalog Interface was being reviewed and approved during the late 1990s, the user paradigm was undergoing a rapid evolution, spurred by advances in web technologies, decreasing hardware costs, increasing online storage capacity, and better network connectivity. Data providers began providing data online via the Web and provided some simple data services to allow the user to access desired data. The old user paradigm of data metadata search and then data order underwent a change. After data metadata search, users expected more sophisticated visualization and analysis capabilities before accessing the data online, possibly invoking some associated data services such as advanced processing, visualization, and subsetting before the data access.

During this same time period, there have been a number of advances in web-based delivery of services that can be used to process Earth Observation (EO) data in meaningful ways. In particular, the OGC has developed (under its Interoperability Program) the Web Map Server (WMS) [3], Web Feature Server (WFS) [4], and Web Coverage Server (WCS) [5] interfaces. Implementations of these interfaces are being developed and instantiated at an increasing rate. This is particularly true for the WMS interfaces, which are the most mature.

The rapid deployment of WMS systems within the EO community and the emerging deployment of WFS and WCS systems has led to an increased need for catalog and discovery mechanisms that will allow end-users to find and employ these services without needing a priori knowledge of their exact service locations.

Additionally, these services themselves contain underlying data content - content that is sometimes exposed via traditional data catalogs and sometimes not. Furthermore, new service types (such as coordinate transformation services) that are essentially content-neutral or content-free are being proposed. The EO community and the OGC have recognized a need for a unified service and data registry concept that provides for publishing and discovery of data, services, or combinations of data and services.

The remainder of this paper tracks the evolution of the OGC service model activities from its beginnings in the Web Mapping Testbed in 1999 through the anticipated developments of the Open GIS Web Services (OWS) Initiative, which will take place in late 2001 and early 2002.

[4] Abstract Syntax Notation - a self-describing data binary data format

IV. PREVIOUS WORK

A. *Web Mapping Testbed - 1999*

The OGC Web Map Server specification was developed in 1999 as part of the OGC Web Mapping Testbed (WMT). A Web Map Server is defined by the specification to be a software server component that renders spatial data in the form of maps output in one of the common image formats such as Portable Network Graphics (PNG). A particular WMS instance advertises its ability to produce maps via a document encoded in XML. This document (called the “Capabilities XML”) contains information about each kind of map that the WMS can produce such as its name, its title, a short abstract, the spatial area it can cover and so on. The level of granularity of this information is on a per "layer" basis where the actual data content of a layer is at the discretion of the WMS provider. Typically a layer could contain world political boundaries or average sea surface temperature over a 12-month period.

Originally the intent of the Catalog specification was to describe interfaces to data catalogs. However during the WMT a catalog was built using the Catalog specification to store metadata about Web Map Server layers. This metadata included the per-layer information described above as well as the Uniform Resource Identifier (URI) [6] that would allow a client program (typically a web browser) to direct a request for a map layer to a particular WMS. Client software could then be employed to search the catalog of WMS metadata for layers and these layers could then be added to a map view being constructed by the client. Thus was born a service catalog.

B. *OGC Interoperability Program - 2000*

Subsequent to the successful completion of the Web Mapping Testbed, the OGC initiated two concurrent test beds (Web Mapping Testbed Phase 2 and Geospatial Fusion Services Testbed Phase 1) under the "Interoperability

Program" banner. These are known individually as WMT-2 and GFST-1, respectively and collectively as IP2000. The results of IP2000 relevant to this discussion were the definition of the Web Feature Server (WFS) and Web Coverage Server (WCS) interfaces (with distinct similarities to the WMS interfaces) as well as three others called the Gazetteer [7], Geocoder [8], and Geoparser [9] interface definitions.

The WFS definition describes interfaces that can return spatial information encoded as Geographic Markup Language (GML) [10]. The interfaces of a WFS are essentially database interfaces including query and transaction interfaces. The WCS definition describes interfaces that can return spatial information encoded as GeoTIFF [11] or HDF-EOS[5]. A WFS is thus geared towards vector-based spatial information (i.e. points, lines, polygons) and a WCS is geared towards gridded spatial information (2, 3, and n-dimensional arrays of numeric information). This is in contrast to the WMS since it can only return pictures of spatial information in the form of maps. Requests to a WFS are encoded either as name/value pairs in a URI or are encoded as XML.

The Gazetteer, Geocoder, and Geoparser interfaces are quite similar to those of a WFS. Requests are typically in the form of a query encoded in XML and results are returned as GML.

The range of new interface types spawned a need for additional service metadata as well as for a coherent model of how to describe the metadata, how to collect it, how to search it and how to exploit it in a system context. The effort to develop such a model became known within IP2000 as the Basic Service Model (BSM) [13]. Several aspects were examined: common elements, terminology, interface types, interface inheritance, service metadata and content metadata. No conclusive specification was developed and it was considered to be a high priority work item for the next round of the OGC Interoperability Program.

V. Current Status

A. Open GIS Service Model

Since the conclusion of IP2000 and the publication of a number of interface specifications and discussion papers (the OGC documents mentioned previously) a group of OGC members has been working to develop the nucleus of a comprehensive service model based on the previous work. At the same time, the general information technology industry is working on similar topics. In particular, the development of the Universal Description and Discovery Interface (UDDI) and the Web Services Description Language (WSDL) are of importance. This work is in preparation for the next phase of the OGC Interoperability Program know as Open GIS Web Services (OWS).

B. Open GIS Web Services (OWS)

The activities of the Open GIS Web Services initiative relevant to this discussion will initially follow two major threads. The first is to develop a common architecture around a General Service Model (GSM). The second, which will coincide to a large degree with the first, is to adapt the existing OGC web service specifications (e.g. WMS, WFS) to the new GSM.

The service model is the overall model governing how to structure and characterize OGC Web Services. The GSM will likely consist of an architecture where individual services have interfaces of known types; the interface types are described in service metadata; the service metadata are available to clients of the service via a "Get Capabilities" request; there are Catalogs or Service Registries that provide queryable access to collections of service metadata; there are services provided by these Catalogs/Registries that assist in maintaining the information contained in the catalog; and the interface types form an inheritance tree of interface properties.

The General Service Model will provide the ability to describe and place into operation services that can be cascaded or chained and services that are loosely coupled to data sources as opposed to being tightly bound to data sources.

The GSM will also be affected by (and hopefully benefit from) wider IT industry efforts such as UDDI, WSDL, and SOAP that provide service models and architectures that OGC can take advantage of and be compatible with.

Service Metadata

Service metadata is that information that describes a given service. There is service type metadata and there is further service instance metadata. For example, a portrayal service will have different service type metadata than a feature service. A feature service could be instantiated on Asian data and another, identical feature service could be instantiated on Antarctica; these services would have the same service type metadata but would have different service instance metadata.

Service metadata should take into account existing metadata standards, in particular those of ISO TC 211, OGC, and those of the particular industry service model used as a foundation.

Capabilities

A given service must provide access to its service type and service instance metadata. Historically within OGC, this is known as the "Capabilities" request and response. The information returned via a Capabilities response is what would be stored in a catalog to provide a searchable repository of services.

[5] HDF-EOS is an API and library of routines that invoke HDF [12] to create standard groups of HDF objects that form HDF-EOS idioms. (From the HDF EOS web site.)

Catalogs

The OGC Catalog Services Interface Specification is the initial source of information regarding construction of a service catalog. It describes the formulation of requests or queries into the holdings of a catalog and describes a mechanism for receiving a response. (Section VI below contains a description of a services catalog prototype.) Specific request and response syntax and semantics are under development to satisfy the needs of the OGC service model. The work has primarily concerned itself with the content of the response, following the pattern developed in the FGDC Geo profile of Z39.50.

Further services in aid of maintaining the coherency and currency of the holdings of a catalog will likely be important additions to the OGC Catalog Services model.

Service Types, Taxonomy, and Interface Inheritance

There are two ways to classify services. One is to place them into a taxonomy - this would be useful to users looking at lists of service names. The other is based on the commonality of the underlying interfaces. For example, a Web Feature Service that allows users to upload features via a transaction interface is a specialization of a Web Feature Service that only allows queries into its existing holdings.

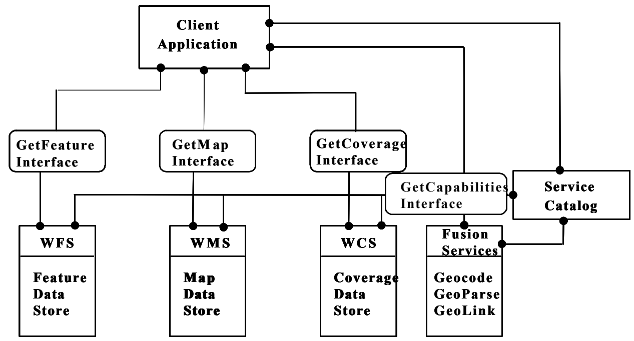
There is a service taxonomy in the Service Architecture portion of the OGC Abstract Specification [14]. The taxonomy presents what can be considered a list of valid search terms that can be used when searching a service catalog.

VI. Services Catalog Prototype

OGC Catalog Services as described above enable clients operating in open distributed environments to quickly and reliably locate and identify desired data, associated services and supporting metadata. This section describes a prototype implementation of an OGC compliant Services Catalog that supports the search and discovery of geo-referenced service metadata describing such Geographic Information services. In addition, the prototype provides a web-based interface for service providers to 'publish' their holdings, as well as providing facilities for the automatic harvesting of this data at pre-specified intervals. In its current form, the prototype supports the interfaces as described in the WMS Interface Specification, with planned future support for other OGC specific service types such as Web Feature Services and Web Coverage Services.

An OGC Web Services Server describes it's capabilities via the GetCapabilities request/response paradigm. In response to a GetCapabilities HTTP request, the Server responds with an XML document that lists all of the services (map layers in the case of a WMS) offered by the server, their coverage areas, spatial reference systems (SRS), available output

Fig. 1. OGC Service Architecture showing interfaces between components



formats and rendering styles, whether the server can answer queries about map contents etc. This XML document is precisely formatted according to a predefined Document Type Definition (DTD). Clients can process this document automatically to select or reject individual services.

In actual use, instead of querying a single OGC Web Services Server, a client might prefer to query a Catalog, which knows about multiple Server instances and can respond to searches for services, satisfying some criteria. The interface context between OGC specific services, and the Services Catalog is depicted in Fig. 1.

Capabilities XML documents from individual servers are harvested by the Service Catalog, processed, and then made available for searching. Since it supports querying via. HTTP (GET or POST), with requests encoded in XML, it supports stateless operations and does not maintain state between search requests. In response to a search request, the Services Catalog returns an XML document, again formatted to a pre-defined response DTD. For efficiency reasons, three profiles of response records have been defined: brief record for a quick look at results, full record for the full capabilities record and a summary record showing an intermediate level of detail. These response record formats, under the control of the client search query, enable clients to take appropriate action without having to parse the entire capabilities document.

Work is currently underway to support other OGC Web Service types as well as formalize the interface for service providers to register their holdings with the Services Catalog. The OGC is looking at some of the newer XML based technologies such as WSDL and UDDI to fill this need.

VII. Current & Future Work & Issues

The Service Model work was started within the OGC Interoperability Program and continues within other OGC programs. Currently, related technologies such as WSDL and UDDI are being analyzed for use in the Service Model. Other relevant technologies developed by other organizations will be identified and analyzed as appropriate. Ideas for integrating data and services are being explored by examining how four services – OGC Data Catalog Service, Web Map Service, Web Feature Service, and Web Coverage Service can be integrated. Service Metadata issues for the four types of services will be studied. Some examples of community defined services that are outside the OGC defined services (e.g. subsetting, data file format translation, etc.) will be studied and ideas for integrating ad-hoc community based services into the Service Model will be examined.

The current prototype implementation consists of a Service Catalog that successfully indexed service metadata from WMS instances. The prototype implementation will be modified to reflect the current state of the Service Model as it evolves.

The Service Model work has generated tremendous interest from government agencies as well as from the commercial companies implementing standards. As more and more geospatial information systems offer online data access and with it, the ability to invoke different data services, the need for a comprehensive service model and data and services catalog becomes more pressing. This initial phase of the Service Model work has produced the following technical challenges and issues :

- How to integrate data and services in the service model,
- Definition of both the system (infrastructure) view and the user view of service metadata,
- How to bundle/link related data and services for the user dynamically,
- How to integrate non-OGC, community defined, data services to the Service Model, and
- How to integrate the Service Model with existing catalog systems.

The initial version of the Service Model is expected to be evolved over the next several years to reflect an increasing complexity of services, tighter integration of services, and more capabilities in the Service Catalog. As the Service Model work continues, additional technical issues are expected to be identified and explored.

REFERENCES

[1] Open GIS Consortium, OpenGIS® Catalog Interface Implementation Specification Revision 1.0, 1999, http://www.opengis.org/techno/specs/99-051.pdf

[2] Bray, Paoli, Sperberg-McQueen, eds., "Extensible Markup Language (XML) 1.0", 2nd edition, W3C Recommendation, October 2000, http://www.w3.org/TR/2000/REC-xml

[3] Open GIS Consortium, OpenGIS® Web Map Server Interfaces Implementation Specification Version 1.1.0, June 2001, http://www.opengis.org/techno/specs/01-047r2.pdf

[4] P. A. Vretanos, Ed., OpenGIS Project Document 01-023: Web Feature Server, 2001, http://www.opengis.org/techno/discussions/01-023.pdf

[5] J. Evans, Ed., OpenGIS Project Document 01-018: Web Coverage Server, 2001, http://www.opengis.org/techno/discussions/01-018.pdf

[6] Berners-Lee, T., Fielding, N., and Masinter, L., "Uniform Resource Identifiers (URI): Generic Syntax", IETF RFC 2396, August 1998, http://www.ietf.org/rfc/rfc2396.txt

[7] R. Atkinson, Ed., OpenGIS Project Document 01-036: GazetteerService Specification, 2001, http://www.opengis.org/techno/discussions/01-036.pdf

[8] S. Margoulies, Ed., OpenGIS Project Document 01-026r1: Geocoder Service Draft Candidate Implementation Specification 0.7.6" 2001, http://www.opengis.org/techno/discussions/01-026r1.pdf

[9] J. Lansing, Ed., OpenGIS Project Document 01-035: Geoparser Service Specification, 2001, http://www.opengis.org/techno/discussions/01-035.pdf

[10] S. Cox, A. Cuthbert, R. Lake, R. Martell, *OpenGIS® Geography Markup Language (GML) Implementation Specification*, February 2001, http://www.opengis.net/gml/01-029/GML2.html

[11] N. Ritter and M. Ruth, GeoTIFF Format Specification GeoTIFF Revision 1.0, Dec. 28, 2000, http://www.remotesensing.org/geotiff/spec/geotiffhome.html

[12] NCSA Scientific Data Technologies HDF Group, HDF Specification and Developers Guide v4.1r4, January 2001, ftp://ftp.ncsa.uiuc.edu/HDF/HDF/Documentation/HDF4.1r4/Spec/

[13] J. de La Beaujardiere, OpenGIS Project Document 01-022r1: Basic Services Model, 2001, http://www.opengis.org/techno/discussions/01-022r1.pdf

[14] The Open GIS Consortium, *The OpenGIS™ Abstract Specification Topic 12: OpenGIS(tm) Service Architecture Version 4.1*, 2001, http://www.opengis.org/public/abstract/01-112.pdf (Section 7.3 - Geospatial Services)